R E P O R T R E S U M E S

ED 010 685 AL 000 043

CHINESE GRAMMARS AND THE COMPUTER AT THE OHIO STATE UNIVERSITY. PRELIMINARY REPORT.

BY- MEYERS, L.F. YANG, J.

PUB DATE MAY 65

EDRS PRICE MF-$0.09 HC-$0.52 13P.

DESCRIPTORS- #GRAMMAR, #TRANSFORMATION THEORY (LANGUAGE), #COMPUTATIONAL LINGUISTICS, PROGRAMING, MANDARIN CHINESE, COMIT, SNOBOL

SAMPLE OUTPUT SENTENCES OF VARIOUS COMIT AND SNOBOL PROGRAMS FOR TESTING A CHINESE GENERATIVE GRAMMAR ARE PRESENTED. THE GRAMMAR CHOSEN FOR EXPERIMENTATION IS A PRELIMINARY VERSION OF A TRANSFORMATIONAL GRAMMAR. ALL OF THE COMIT PROGRAMS AND ONE OF THE SNOBOL PROGRAMS USE A LINEARIZED REPRESENTATION OF TREE STRUCTURES, WITH ADDITIONAL NUMERICAL SUBSCRIPTS FOR EASIER IDENTIFICATION OF CORRESPONDING PARENTHESES. A SNOBOL PROGRAM USES INSTEAD A LIST OF NODE NUMBERS, TOGETHER WITH IMMEDIATE ANCESTORS AND DESCENDANTS. A SET OF SUBROUTINES NOW BEING WRITTEN IN COMIT USES MODIFIED POLISH NOTATION, IN WHICH A SUBSCRIPT GIVES THE NUMBER OF NODES IMMEDIATELY DOMINATED BY A GIVEN ONE. ONE SUCCESSFUL COMIT PROGRAM FOR THE PHRASE STRUCTURE OF THE GRAMMAR IS ALSO EXTENDED TO A FEW TRANSFORMATION RULES, WHICH INVOLVE EMBEDDING. THIS ARTICLE APPEARS IN THE OHIO STATE UNIVERSITY RESEARCH FOUNDATION PROJECT ON LINGUISTIC ANALYSIS, REPORT NUMBER 10. (IT)

ED010685

RF Project 1685-3

THE OHIO STATE UNIVERSITY

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

RESEARCH FOUNDATION

1314 KINNEAR ROAD COLUMBUS, OHIO 43212

Project on Linguistic Analysis
Report No. 10

CHINESE GRAMMARS AND THE COMPUTER AT THE OHIO STATE UNIVERSITY

Leroy F. Meyers and James Yang
Divisions of Mathematics and
Electrical Engineering

AL 000 043

May 1965
NATIONAL SCIENCE FOUNDATION
Grant No. GN-174

PRECEDING PAGES MISSING

BIBLIOGRAPHY

Annear, S., "English and Mandarin Chinese: The Comparative Construction," unpublished term paper.

_____, "A Case of Pronominalization in the Modifying Clause Construction in Mandarin," unpublished term paper.

_____, and L. Liu, "A Fragment of Mandarin Grammar: BA and BEI", POLA Report No. 7, pp. 63-75.

Chomsky, A. N., Aspects of the Theory of Syntax, in press.

Fillmore, C. J., "Entailment Rules in a Semantic Theory," paper read before the 1964 International Symposium on Signs and Systems of Language.

Hashimoto, A. Y., "Resultative Verbs and Other Problems," POLA Report No. 8.

_____, "Adjectives in Mandarin," unpublished term paper.

_____, "YOU," unpublished term paper.

_____, "Modifying Clause Construction in Mandarin," unpublished report.

Shih, M., "The Negatives in Mandarin," unpublished term paper.

_____, "Relationship between Yes-No Questions and Answers," unpublished term paper.

Wang, W. S-Y., "Some Syntactic Rules in Mandarin," POLA Report No. 3r, pp. 32-53; also Proceedings of the Ninth International Congress of Linguists, pp. 390-401.

_____, "Two Aspect Markers in Mandarin," POLA Report No. 8 and to appear in Language.

_____, "Conjoining and Deletion in Mandarin Syntax," paper read at the summer meeting of LSA, 1964.

AL 000 043

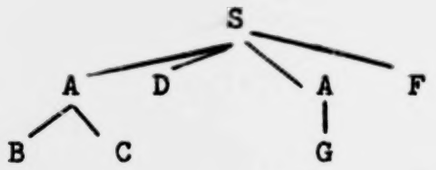
CHINESE GRAMMARS AND THE COMPUTER AT THE OHIO STATE UNIVERSITY

(Preliminary Report)

L. F. Meyers and J. Yang

Planning for testing a Chinese generative grammar on a high-speed digital computer began in May, 1964. The grammar chosen for experimentation is a preliminary version of a transformational grammar prepared by Anne Yue Hashimoto. James H. Yang and Leroy F. Meyers have done the programming for it, using both COMIT and SNOBOL.

Since transformational grammars require manipulation of tree structures, an appropriate machine representation of trees is needed. For example, the tree:



can be linearized, using parentheses, as:

$$(_S\ (_A\ B\ C\ )_A\ D\ (_A\ G\ )_A\ F\ )_S\ .$$

All of the COMIT programs and one of the SNOBOL programs have used this representation, with additional numerical subscripts for easier identification of corresponding parentheses, as in:

$$(_{1,S}\ (_{2,A}\ B\ C\ )_{2,A}\ D\ (_{3,A}\ G\ )_{3,A}\ F\ )_{1,S}\ .$$

A SNOBOL program uses instead a list of node numbers, together with immediate ancestors and descendants:

1. 0 S.3 4. 2 3 4 5
2. 1 A.2 2. 6 7
3. 1 D.0 0.
4. 1 A.5 1. 8
5. 1 F.0 0.
6. 2 B.0 0.
7. 2 C.0 0.
8. 4 G.0 0.

For example, node 2 is dominated by node 1, is labeled A, and uses subrule number 2 to dominate the two nodes numbered 6 and 7.

A set of subroutines now being written in COMIT uses modified Polish notation, in which a subscript tells how many nodes are immediately dominated by a given one:

$$S_4\ A_2\ B\ C\ D\ A_1\ G\ F\ .$$

At present, Yang's successful COMIT program for the phrase structure portion of the grammar is being extended to take care of the first few transformation rules, which involve embedding. One difficulty with programming the embedding rules (besides their complexity) is that it is seldom possible to generate two sentences at random such that one can be embedded into the other.

Some sample output sentences are given on the next few pages. These are based on various slight modifications of the rules in Mrs. Hashimoto's grammar. The corresponding trees are also illustrated.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Sample Output of Yang Program 2 (Constituent Structure only; COMIT)

Sentence (001) ZHE4 YI1 GE4 GOU3 MOD DE XIANG3 WO3MEN

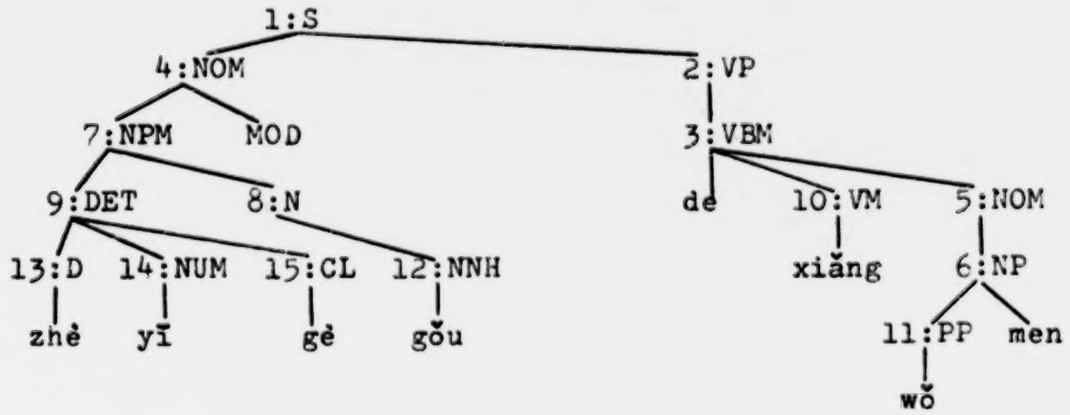
Tree description

```
- + *( / .1 , S + *( / .4 , NOM + *( / .7 , NPM + *( / .9 ,
DET + *( / .13 , D + -ZHE*4+ *) / .13 , D + *( / .14 , NUM
+ -YI*1 + *) / .14 , NUM + *( / .15 , CL + -GE*4 + *) / .15
, CL+*) / .9 , DET + *( / .8 , N + *( / .12 , NNH + -GOU*3+
*) / .12 , NNH + *) / .8 , N + *) / .7 , NPM + -MOD + *) /
.4 , NOM + *( / .2 , VP + *( / .3 , VBM + -DE + *( / .10 ,
VM + -XIANG*3 + *) / .10 , VM + *( / .5 , NOM + *( / .6 , NP
+ *( / .11 , PP + -WO*3 + *) / .11 , PP + MEN + *) / .6 , NP
+ *) / .5 , NOM + *) / .3 , VBM + *) / .2 , VP + *) / .1 ,
S +
```

Rules and subrules used

```
(001) S A + VP D + VBM D + NOM B + NOM A + NP C + NPM C + N C
+ DET C + VM + PP + NNH + D + NUM + CL
```

Tree



Sample Output of Meyers Program 1 (Constituent Structure only: COMIT)

Rules used

```
S + *= + NOM + VP +
NOM + *= + NP +
NP + *= + PP + MEN +
PP + *= + -WO*3 +
VP + *= + VBM +
VBM + *= + -DE + VM + NOM +
VM + *= + -AI*4 +
NOM + *= + NP +
NP + *= + DET + N +
DET + *= + NUM + CL +
NUM + *= + -ER*4 +
CL + *= + ZHANG*1 +
N + *= + NNH +
NNH + *= -CHE*1+
```

Sentence

```
001.  WO3 MEN  DE  AI4  ER4 ZHANG1 CHE1.
```

Tree description

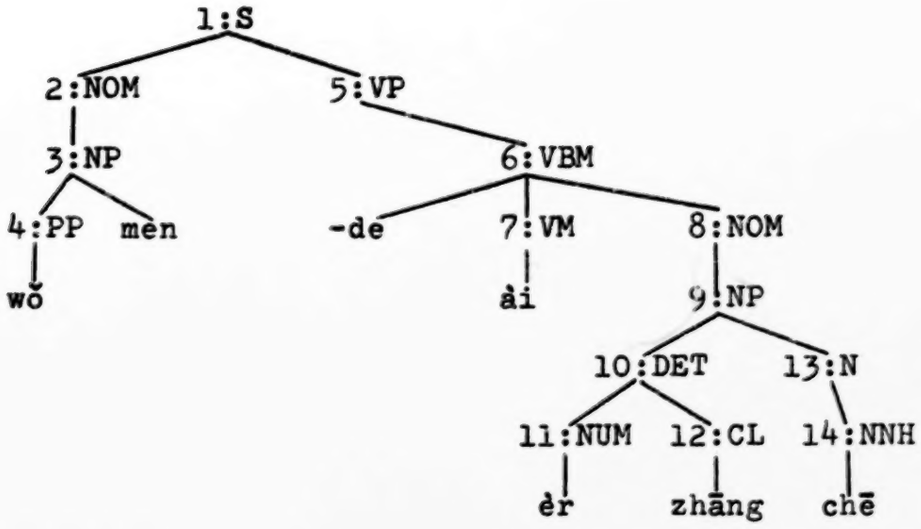
```
-*0*0*1. + *(S / .1 + *(NOM / .2 + *(NP / .3 + *(PP / .4 +
-WO*3 + PP*) / .4 + MEN + NP*) / .3 + NOM*) / .2 + *(VP / .
+ *(VBM / .6 + -DE + *(VM / .7 + -AI*4 + VM*) / .7 + *(NOM
/ .8 + *(NP / .9 + *(DET / .1) + *(NUM / .11 + -ER*4 + NUM*
/ .11 + *(CL / .12 + ZHANG*1 + CL*) / .12 + DET*) / .10 + *
/ .13 + *(NNH / .14 + -CHE*1+ NNH*) / .14 + N*) / .13 + NP*
/ .9 + NOM*) / .8 + VBM*) / .6 + VP*) / .5 + S*) / .1 +
```

Rules

```
S NOM NP PP VP VBM NOM NP DET NUM CL N NNH
```

Sample Output of Meyers Program 1 (Constituent Structure only: COMIT)

Tree



Sample Output from Meyers Program 2 (Constituent Structure; SNOBOL)

Rules used

S = NOM VP F
NOM = NP
NP = PP
PP = NI3
VP = VBNST
VBNST = ASP VQUO C
ASP = LE
VQUO = SHUO1
F = MA (Hello, Newton!)

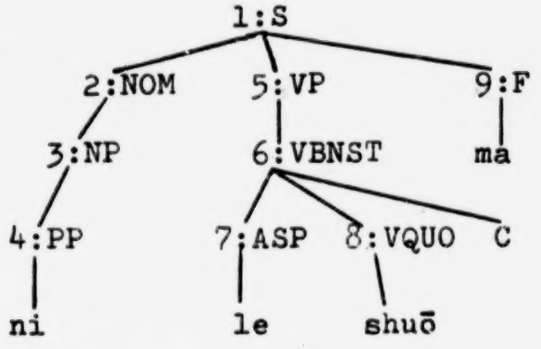
Sentence

1. NI3 LE SHUO1 C MA .

Tree description

1. (.1.S (.2.NOM (.3.NP (.4.PP NI3).4.PP).3.NP).2.NOM (.5.VP (.6.VBNST (.7.ASP LE).7.ASP (.8.VQUO SHUO1).8.VQUO C).6.VBNST).5.VP (19.F MA).9.F).1.S

Tree



Sample Output of Yang Program 7 (Constituent Structure and the First Transformational Rule; COMIT)

Sentence

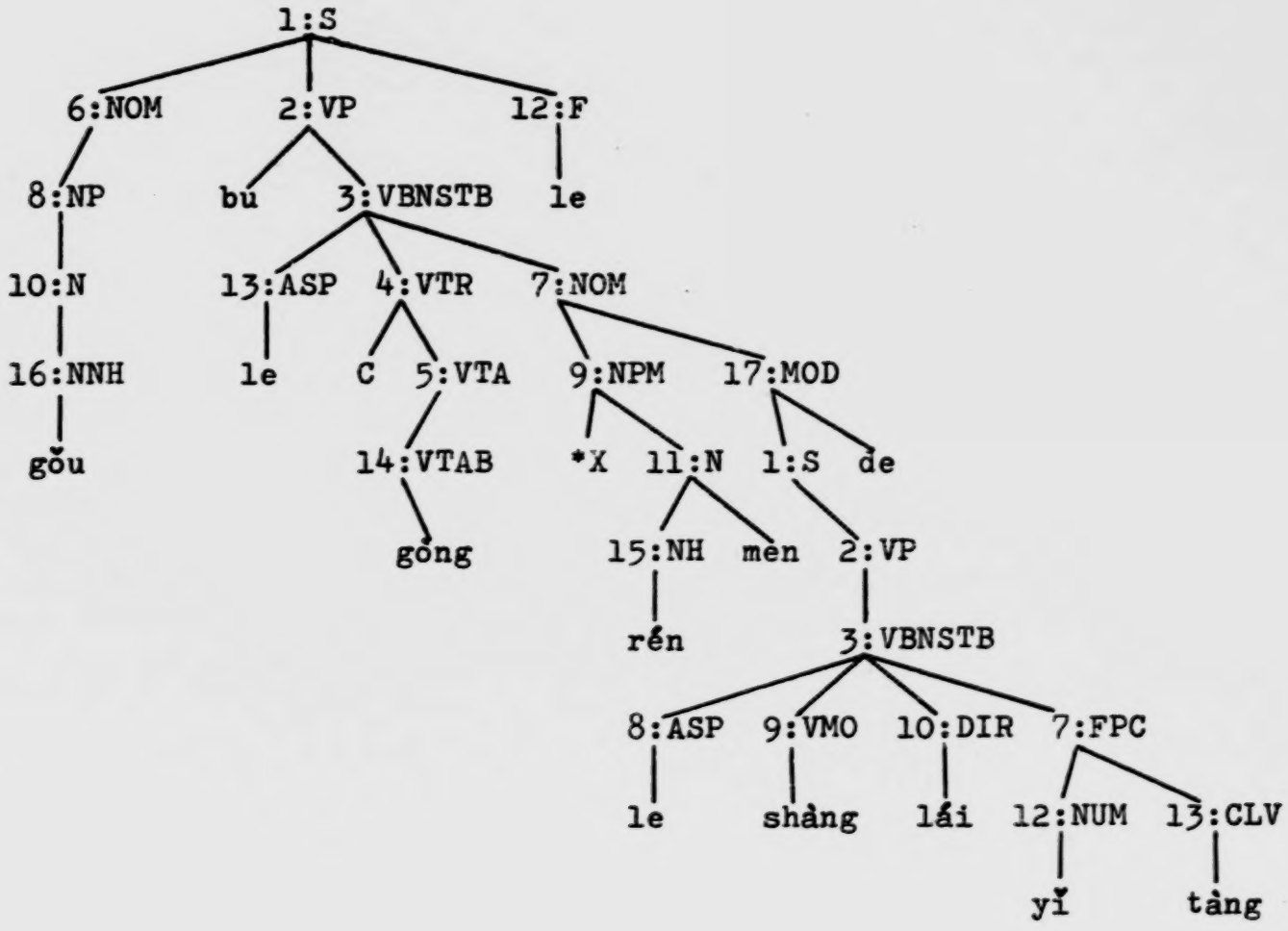
OUTPUT-OF-ETR1 (((((GOU3))))(BU((LE)(C((GONG4)))((X((
REN2)MEN))(((((LE)(SHANG4)(LAI2)((YI1)(TANG4))))) DE))))(LE))

Tree description

(/ .1 , S + *(/ .6 , NOM + *(/ .8 , NP + *(/ .10 , N +
*(/ .16 , NNH + -GOU*3+ *) / .16 , NNH + *) / .10 , N + *)
/ .8 , NP + *) / .6 , NOM + *(/ .2 , VP + -BU + *(/ .3 ,
VBNSTB + *(/ .13 , ASP + -LE + *) / .13 , ASP + *(/ .4 ,
VTR + -C + *(/ .5 , VTA + *(/ .14 , VTAB + -GONG*4 + *) /
.14 , VTAB + *) / .5 , VTA + *) / .4 , VTR + *(/ .7 , NOM
+ *(/ .9 , NPM + *) + *(/ .11 , N + *(/ .15 , NH + -REN*2+
*) / .15 , NH + MEN + *) / .11 , N + *) / .9 , NPM + *(/ .17
, MOD + *(/ .1 , S + *(/ .2 , VP + *(/ .3 , VBNSTB + *(
/ .8 , ASP + -LE + *) / .8 , ASP + *(/ .9 , VMO + -SHANG*4
+ *) / .9 , VMO + *(/ .10 , DIR + -LAI*2+ *) / .10 , DIR +
*(/ .7 , FPC + *(/ .12 , NUM + -YI*1 + *) / .12 , NUM + *(
/ .13 , CLV + -TANG*4 + *) / .13 , CLV + *) / .7 , FPC + *)
/ .3 , VBNSTB + *) / .2 , VP + *) / .1 , S + -DE + *) / .17
, MOD + *) / .7 , NOM + *) / .3 , VBNSTB + *) / .2 , VP + *(
/ .12 , F + -LE + *) / .12 , F + *) / .1 , S +

Sample Output of Yang Program 7 (Constituent Structure and the First Transformational Rule; COMIT)

Tree



Sample Output of Meyers Program 3 (Constituent Structure only; SNOBOL)

Tree description

```
 1.  0 S.1 2. 2 3
 2.  1 NOM.0 1. 4
 3.  1 VP.5 1. 5
 4.  2 NP.1 2. 6 7
 5.  3 VBM.2 3. 8 9 10
 6.  4 PP.1 1. 11
 7.  4 MEN.0 0.
 8.  5 EMP.1 1. 12
 9.  5 VM.1 1. 13
10.  5 NOM.0 1. 14
11.  6 NI3.0 0.
12.  8 HAO3.0 0.
13.  9 AI4.0 0.
14.  10 NP.0 2. 15 16
15.  14 DET.2 2. 17 18
16.  14 N.1 1. 19
17.  15 D.1 1. 20
18.  15 CL.1 1. 21
19.  16 NNH.1 1. 22
20.  17 NA4.0 0.
21.  18 ZHANG1.0 0.
22.  19 CHE1.0 0.
```

Sentence

```
1.  NI3 MEN HAO3 AI4 NA4 ZHANG1 CHE1.
```

Sample Output of Meyers Program 3 (Constituent Structure only; SNOBOL)

Tree